(2)

SECURITY CLASSIFICATION OF THIS PAGE

AD-A264 342

REPORT DOCUMENTATION PAGE

DTIC ELECTE MAY 14 1993 S C D

1b. RESTRICTIVE MARKINGS

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
F49620-92-J-0169

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 93 0287

6a. NAME OF PERFORMING ORGANIZATION
Yale University

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NL

6c. ADDRESS (City, State and ZIP Code)
12 Prospect Place
New Haven, CT 06511

7b. ADDRESS (City, State and ZIP Code)
110 Duncan Ave, Suite B115
Bolling AFB DC 20332/0001

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR/NL

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F49620-92-J-0169

8c. ADDRESS (City, State and ZIP Code)
Building 410
Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313/AS	2313/BS	

11. TITLE (Include Security Classification)
Representation of Shape in Object Recognition and Long-Term Visual Memory

12. PERSONAL AUTHOR(S)
Tarr, Michael J.

13a. TYPE OF REPORT
Annual

13b. TIME COVERED
FROM 1/15/92 TO 1/14/93

14. DATE OF REPORT (Yr., Mo., Day)
93/2/11

15. PAGE COUNT
30

16. SUPPLEMENTARY NOTATION

93 5 13 015

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Object Representation, Object Recognition
Visual Cognition

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A variety of studies examining the mechanisms and representations underlying human object recognition have been conducted. One track has investigated the role of view-based object representations in perception and recognition. Results indicate that certain classes of viewpoint-dependent features may be used to define boundaries between characteristic views of objects. A second track has investigated the interaction between orientation-dependent and orientation-independent recognition mechanisms. Results here indicate that humans learn both object-based, orientation-independent and view-based, orientation-dependent representations regardless of the initial learning context. Other results indicate that task conditions mediate whether structural descriptions or episodic representations of objects are used in performing an implicit memory task. Finally, a third track has investigated the nature of spatial relations between objects, as well as the relationship between perceptual and lexical representations of spatial relations. Results indicate that spatial prepositions (e.g., "above", "left") encode the relationship between figural and reference objects as a gradient that decreases with distance from the qualitative or veridical position. Moreover, results indicate that this may in part be a lexical effect, in that stronger qualitative effects are found when subjects have lexically encoded the relationship — although further results indicate that qualitative gradients are present in purely perceptual judgments.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. John Tangney

22b. TELEPHONE NUMBER (Include Area Code)
(202) 767-5021

93-10718

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

17 FEB 1993

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR Grant No. F49620-92-J-0169

REPRESENTATIONS OF SHAPE IN OBJECT RECOGNITION AND LONG-TERM VISUAL MEMORY

Michael J. Tarr
Yale University
Department of Psychology
PO Box 11A Yale Station
New Haven, CT 06520-7447

February 10, 1993

Annual Technical Report for Period 15 January 1992 - 14 January 1993

Distribution Statement

Prepared for

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Building 410
Bolling AFB, DC 20332-6448

ANNUAL TECHNICAL REPORT

REPRESENTATIONS OF SHAPE IN OBJECT RECOGNITION AND LONG-TERM VISUAL MEMORY

Michael J. Tarr
Yale University

AFOSR Grant #F49620-92-J-0169

ABSTRACT

A variety of studies examining the mechanisms and representations underlying human object recognition have been conducted. One track has investigated the role of view-based object representations in perception and recognition. Results indicate that certain classes of viewpoint-dependent features may be used to define boundaries between characteristic views of objects. A second track has investigated the interaction between orientation-dependent and orientation-independent recognition mechanisms. Results here indicate that humans learn both object-based, orientation-independent and view-based, orientation-dependent representations regardless of the initial learning context. Other results indicate that task conditions mediate whether structural descriptions or episodic representations of objects are used in performing an implicit memory task. Finally, a third track has investigated the nature of spatial relations between objects, as well as the relationship between perceptual and lexical representations of spatial relations. Results indicate that spatial prepositions (e.g., "above", "left") encode the relationship between figural and reference objects as a gradient that decreases with distance from the qualitative or veridical position. Moreover, results indicate that this may in part be a lexical effect, in that stronger qualitative effects are found when subjects have lexically encoded the relationship — although further results indicate that qualitative gradients are present in purely perceptual judgments.

STATEMENT OF WORK

The objective of this project is to investigate the mechanisms and underlying representations implicated in human visual cognition, in particular, in object recognition, mental imagery, and visual problem solving. While several different approaches have been taken, the fundamental assumption behind all of the studies conducted to date is that visual object recognition is accomplished by matching an input shape to an object representation stored in long-term visual memory. In accordance with earlier work by the PI, as well as many others (Farah, 1992; Jolicoeur, 1990), there is evidence that at least two distinct mechanisms are available for recognition: one an orientation-independent mechanism utilizing object-based representations; and another an orientation-dependent mechanism utilizing view-based representations. Consistent with this hypothesis, research questions have been formulated to examine: the capacities and limitations of these mechanisms in situations where the operation of a given mechanism is unambiguous; the interaction between these mechanisms under more general conditions (such as those usually encountered in the "real-world"); and, finally, the format of the object representations upon which these mechanisms depend.

One important aspect of the approach is the application of computational theory whenever possible. This includes not only precise formulations of problems in terms of constraints and algorithms (Marr, 1982), but some adherence to advances made in computer vision — a discipline not overly concerned with modeling human performance, but offering a wide range of formal solutions to problems in vision encountered by both humans and machines. Such an approach may have some advantages over more

traditional approaches to studying human cognition. First, models taken from computer vision tend to be specified in detail and have frequently been implemented and tested for flaws, idiosyncrasies, and limitations — often using real-world images. Second, application of computational and formal models to human vision may reveal heretofore unrealized algorithms and solutions that may be used to refine artificial vision systems. Third, the development of accurate simulations of psychological processes may facilitated by the introduction of mechanisms and representations where some thought has been given to their computational properties. Finally, computational models provide a clear framework for characterizing and understanding the properties of psychological mechanisms.

Examples of this approach may be seen in many of the studies discussed in the next section. First, in Section 1, a formal analysis of object geometry has been applied to the problem of how humans define the boundaries between views of an object. Results here suggest not only how the theory of human performance should be constrained in future studies, but how computer vision implementations may be rendered more tractable. Moreover, the findings of this study may provide an indication of the types of features used in a variety of perceptual tasks — from judging the orientation of a shape, planning mental transformations, to recovering 3D part descriptions. Next, in Sections 4 and 5 there has been an effort to map out and define the stimulus configurations and task contexts that apparently determine the appropriateness of each recognition mechanism. Following Tarr and Pinker's (1990) hypothesis that objects that may be differentiated by linear orderings of parts, as well as non-directional inboard/outboard relationships, may be recognized through orientation-independent mechanisms, but that objects that may be differentiated only by two-dimensional or greater relationships require orientation-dependent mechanisms we have developed a new stimulus set for the purpose of testing this distinction. Accordingly, recognition performance, and thus the requisite recognition mechanism, may be predicted by which stimulus items are included in the recognition set. Moreover, recognition performance should be malleable, depending not only on prior experience, but on recognition task context, as well as the nature of the task itself. Finally, the work described in Section 6 addresses additional aspects of object representations. Tarr (submitted) has argued that object representations may be characterized along four independent dimensions: (1) the frame of reference; (2) the number of spatial dimensions; (3) the description of parts; and (4) the encoding of spatial relations. Most of the work until this point has addressed only the first issue; the work described in Section 6 attempts to begin to place constraints on the commonly held intuitive conception of the fourth issue — that our knowledge of explicitly coded spatial relations corresponds to the spatial prepositions we linguistically encode. Indeed, to this point, little work has been done to describe such relations in anything other than linguistic terms (Biederman, 1987, for example).

STATUS

1. Viewpoint-Dependent Features in Object Representation

Tarr and Pinker (1989) proposed the *Multiple-Views-Plus-Transformation* theory of object recognition. The foundation of this theory is that objects are represented in visual memory as a collection of linked viewer-centered representations (e.g., "views") — each view depicting the characteristic appearance of an object from a restricted range of orientations. Because this conception of object representation is consistent with recent work in a diverse range of disciplines, there is reason to believe that the view-based

approach holds great promise for understanding biological, and in particular human, object representation and recognition.

One of the most important aspects of any view-based approach is the mechanism by which the perceptual system determines which views of a given object should be retained. While typical computer-based recognition system rely on 3D object models that are specified *a priori* by the designer, an active, learning perceiver, such as a human being, must acquire views of objects over space and time without the benefit of preprogrammed object models. Studies by the PI on how humans learn novel views suggest that three distinct factors play a role: familiarity with an object in a given view; familiarity with visually similar objects ("cohorts") in a given view; and the geometry of the visible surfaces of the object in a given view. The first two factors may be characterized purely by probabilistic measures, e.g., how often an object or one of its cohorts is observed in a particular view. However, the surface geometry of most natural objects is highly complex, therefore intuition or other *ad hoc* methods do not provide good theories of when an object shifts from one characteristic or unique view to another. It is here that formal techniques may be applied most effectively.

Koenderink (1987) has proposed geometric methods for determining the topologically distinct views of an object. Starting with a 3D model, this decomposition, referred to as an *aspect graph*, provides a complete representation of every unique view of an object. More specifically, the space of viewpoints can be partitioned into maximal regions wherein the structure of the line drawing defined by image intensity discontinuities (edges) is identical; the regions are delineated by visual events where the structure changes. The structure (topology) of the line drawing is defined by the relationship of feature points such t-junctions, vertices, contour terminations (cusps), inflections, etc. and the smooth contours connecting them. Thus, the object's appearance is qualitatively similar for all orientations within a region; a qualitative change occurs when the orientation crosses a visual event boundary. Importantly, results from singularity and catastrophe theory indicate that there is a relatively small catalogue of visual events and consequently only a small number of ways that the image structure can change.

If humans do use view-based representations (even if such mechanisms are used only for particular tasks), then a principled, geometric decomposition of the view space of objects is necessary for organizing viewer-centered information. Furthermore, because representations of objects are not specified *a priori*, we must learn them as we explore our environment — presumably using image features similar to those specified by computational theory. Consequently, formal descriptions of object geometry, including but not limited to current aspect graph methods, offer the experimental psychologist a principled means for both manipulating the orientation of objects across surface geometry and analyzing human recognition performance and perceptual behavior.

While knowledge of the image features that define visual events is helpful in understanding object structure, it is insufficient for utilizing aspect graph models to study human shape representation. One must also have the means for decomposing actual objects into their characteristic views. This requirement has presented a major obstacle in employing such models in behavioral studies. Crucially, new results have demonstrated that Koenderink's theory is computationally tractable, and it has since enjoyed increasing popularity in the computer-vision community. Even still, the majority of work has focused polyhedral objects; only recently have there been techniques for

computing the complete aspect graphs of a variety of objects based on a combination of catastrophe theory, algebraic geometry, and robust numerical methods.

In order to assess the viability of this framework, the PI, along with David J. Kriegman (Assistant Professor of Electrical Engineering, Yale University), has initiated several studies to capitalize on these computational methods in psychophysical studies. We have begun by conducting a series of experiments to investigate whether humans are indeed sensitive to the features used in determining the topologically distinct views of an aspect graph. The subjects' task was to judge whether two consecutively presented images of the same smoothly curved object (rendered with occluding contours or Phong shaded) were displayed at the same or at different orientations (generated by rotations in depth). Performance was assessed by measuring their accuracy in detecting an orientation difference between the two images. As accuracy increases subjects are demonstrating an increased ability for discriminating a change in view. When one compares the locations of the visual events as predicted by the computational theory (Figures 1, 2, and 3) — that is the orientations where the aspect graph makes the transition from one view to another (denoted by the vertical bars) — to the percent correct function, it is clear that accuracy in discriminating orientations does increase when images cross a visual event. In general, image pairs adjacent separated by visual events exhibited large increases in sensitivity. The second study also included a condition in which the two sequentially presented images appeared at different random screen positions in order to ensure that subjects were not simply attending to a single region of the display — results in this condition were consistent with the fixed position conditions (Figure 3), suggesting that this was not a problem.

The first study used a relatively simple solid of revolution, a torus, with results generally adhering to the predictions of the computational theory (Figure 1). Interestingly, in the subsequent study, using a more complex piecewise smooth object, a bell, we observed that humans are insensitive to some visual events (Figures 2 and3). For example, the visual event at 72.2°, a transition from a pair of features (a t-junction and cusp) to a 3-tangent junction, is actually centered within a region where subjects are very poor at discriminating between views. Similar, but somewhat less dramatic, effects were found at two other visual events (107.8° and 117.8°). Why might humans ignore these particular events? First, the changes in the configurations of image features at these orientations are quite subtle. This intuition is confirmed by the experimental results, but could not have been predicted by the present computational theory. These results may be significant, not only because of what they tell us about human perception, but because they may help to refine machine recognition systems based on view-based approaches.

Second, because of the relative size of some configurations of features, changes occur only gradually over wide shifts in orientation. For example, two areas of great interest are scale and resolution. Current computational derivations of aspect graphs give equal weight to features at *all scales*, creating the potential for an explosive number of unique views for a single complex object (this has been one strong criticism of the aspect graph approach). Thus, in order to utilize aspect graph like structures for object representation, any resource limited perceptual system acquiring information over a finite amount of time must reduce the number of potentially different views per an object. In light of these constraints, as well as general parsimony, humans presumably disregard features at some scales, thereby significantly reducing the number of retained views. Additionally, an understanding of the specific types and configurations of features that are less essential to human performance and under what circumstances they are

considered irrelevant may be used to refine aspect graph construction algorithms by indicating which ones are most pertinent for recognition.

In summary, these initial results are highly promising in that they suggest one method for addressing the question of how to empirically define what constitutes a unique view of an object. Moreover, because they rely on theory derived from computer vision, they may ultimately provide further constraints on developing successful artificial vision systems. Additionally, the fact that humans are apparently sensitive to these viewpoint-dependent features, as well as the way in which they change in configuration with changes in orientation, suggests that such features may be generally relevant to the processes by which humans extract the orientations of shapes, plan and execute mental transformations, and, possibly recover invariant descriptions of parts.

2. Representation of Natural Objects

Most studies of recognition to this point have involved either novel objects composed of either lines or cubes, line drawings of natural objects from a single viewpoint, or simple line drawings of 3D objects. Over the past year we have developed a set of stimuli based on shaded 3D models that may be displayed and illuminated from any chosen orientation and with or without color information. Most of the objects we have developed are either human artifacts or simple lifeforms (e.g., fruit). However, this collection has recently been supplemented by the donation of approximately 30 models of animals from *Viewpoint Animation Engineering, Inc.* With the development of this stimulus set, we will now be able to address many questions concerning the properties of view-based and object-based representations for natural objects. We have begun by collecting baseline data from over 30 subjects on the preferred or "canonical" orientation for each object, as well as subjects' ratings of preference in comparison to other views of the same object (Table 1). This is similar to the set of line drawings developed by Snodgrass and Vanderwert (1980), and currently used as the standard stimuli in many recognition studies (Jolicoeur, 1985, for instance).

We have now initiated several studies with these objects, including an examination of some of the assumptions of Biederman's (1987) "Recognition-by-Components" theory of object recognition. For instance, Biederman has demonstrated invariance across size, mirror reflection, and orientation, suggesting that objects are represented as orientation-independent object-based models. Using a variety of implicit memory priming techniques, we are assessing the generality of these findings. Preliminary results suggest that there are some conditions under which natural objects are apparently stored in a specific view, lending support to the hypothesis that object representations may be view-based, but with a concurrent object-based description. Numerous studies using these stimuli are planned for the coming year.

3. Object Representations Underlying Memory Systems

One of the least explored aspects of visual cognition is the fact that the object representations implicated in object recognition and mental imagery are presumably manifestations of visual long-term memory. As such, they should be subject to the same phenomena and disassociations observed for semantic memory. For instance, Schacter (Schacter, 1987) has demonstrated a strong disassociation between implicit memory, e.g., facilitation due to prior exposure without conscious awareness, and explicit memory, e.g., the conscious recall of an event, item, or episode. Interestingly, Schacter and Cooper (Cooper & Schacter, 1992; Schacter, Cooper, Delaney, Peterson, & Tharan, 1991; Schacter, Cooper, & Delany, 1990) have recently demonstrated a

disassociation in the object representations apparently underlying implicit and explicit visual memory. Specifically, they have shown that impossible objects — 2D line drawings that can not exist as 3D objects except as an accident of viewpoint — mediate priming in explicit, but not in implicit memory tasks. In contrast, possible objects mediate priming in both types of tasks. From this and related results they conclude that there is a disassociation in object representations, with 3D "structural descriptions" underlying implicit memory and 2D "episodic" representations underlying explicit memory.

However, while these results are consistent with the hypothesis that object recognition and representation is a product of both view-based and object-based representations, it is inconsistent with the claim that stimulus context and task constraints determine which mechanisms and representations come into play. Indeed, if Schacter and Cooper are correct, current theories of object recognition will have to be reconsidered in light of the constraints imposed by memory subsystems. It is also possible that Schacter and Cooper's results are an artifact of the particular tasks they have chosen to use — in particular, implicit memory was always assessed by the use of an "object decision" task in which subjects decide whether an object is possible or impossible. The idiosyncratic property of this task is that it inherently *requires* that the 3D structure of the object be extracted in order to perform the task. Thus, it is not surprising that evidence for 3D structural descriptions was found in this task — what is missing is an implicit task that does not intrinsically rely on such representations of objects.

To test whether the lack of priming (e.g., implicit memory) for impossible objects reflects a genuine disassociation between object representations, we have developed a new measure of implicit memory. This task involves overlaying a common natural object with a novel object, either possible or impossible, and degrading the display by randomly deleting a set percentage of cells from a 10 x 10 grid covering the object pair (Figure 4). Each object pair is then displayed as series of degraded images, progressing from highly degraded (many grid points deleted) to marginally degraded (few grid points deleted). The subjects' task is simply to name the common object at the maximal level of degradation possible. The crucial manipulation is that some of the novel objects have been previously studied in perceptual discrimination task, for instance, judging whether each object is point to the left or to the right. Presumably, if subjects have represented these studied objects in memory, they should be somewhat better at segregating them from the common objects, and therefore should be better able to name such objects at higher levels of degradation.

In the first experiment using this technique, we have found promising results. Using only possible novel objects, it was demonstrated that subjects are faster (in terms of level of degradation) to name common objects obscured by studied novel objects as compared to non-studied novel objects (mean level to naming, lower is greater degradation):

Studied	Non-Studied	
4.8857	5.1829	$F(1,34) = 31.4$, $p < .001$

We are now extending the same technique to possible and impossible objects in a direct test of Schacter and Cooper's claims of disassociation. Importantly, if we can demonstrate that impossible objects can be primed in an implicit memory task, then this suggests that apparent differences between structural descriptions and episodic representations manifest themselves according to task demands and possibly training

context, but not as a broad-based distinction between memory subsystems. Moreover, given this new technique for assessing implicit memory for objects — a task in which the primed item is simply to be ignored and therefore is not subject to certain task limitations (for example, requiring that novel objects be assigned names) — we may be better able to examine the format of object representations, as well as their role in long-term visual memory.

4. *Contextual Effects on Object Learning*

Current theories of object recognition have posited two distinct modes of representation. However, it is still unclear as to what conditions determine how perceptual mechanisms apply such representations under different contexts in learning and recognition. Specifically, several researchers have argued that the default is object-based orientation-independent representations sufficient for "basic-level" categorization (Biederman, 1987; Corballis, 1988). Alternatively, it is possible that both object-based and view-based representations are stored regardless of context — each being invoked according to the specific task at a given time. Consistent with this alternative is the basic assumption of the Multiple-Views-Plus-Transformation theory, that objects are represented routinely as a collection of viewpoint-specific "snapshots" or views and that this view-based representation may be invoked when the recognition discrimination necessitates contrasting objects along more than two dimensions.

To investigate this question, a paradigm has been developed in which subjects initially learn and recognize a set of stimuli that may be differentiated by part orderings along a single linear dimension. Tarr and Pinker (1990) have shown that such shapes are immediately and consistently recognized independently of their orientation. Consequentially, throughout subjects' exposure to these shapes in a select set of orientations, there are only marginal effects of orientation, e.g., relatively small slopes (Figure 5), again indicating that orientation-independent recognition mechanisms are apparently used. However, while performance was orientation-dependent throughout this practice phase, suggesting that subjects encoded object-based descriptions, it was hypothesized that subjects are also learning the appearance of each shape in each familiar orientation, e.g., that view-based representations are stored concurrently regardless of learning context. To test this possibility, after extensive practice, subjects are taught three new shapes that vary from the original three along an additional dimension — therefore, to differentiate such shapes from the original shapes, as well as each other, subjects must attend to at least two orthogonal dimensions along which parts may be located. Tarr and Pinker have shown that such contrasts lead to the use of orientation-dependent recognition mechanisms utilizing view-based representations — as a result, it was expected that effects of orientation not previously observed for the recognition of the familiar shapes would manifest themselves with the introduction of these new shapes. Moreover, if subjects have learned view-based representations at familiar orientations during the practice phase, then effects of orientation over a wider range of orientations should vary systematically with distance from the nearest familiar orientation.

As illustrated in Figure 6, subjects apparently have derived more than object-based representations during practice. In particular, their response times for the familiar or "old" shapes generally increase with increasing separation from a practice orientation (the vertical bars in the top panel) — the hallmark of an orientation-dependent recognition mechanism utilizing orientation-specific representations (Jolicoeur & Milliken, 1989; Tarr, 1990, have run studies addressing the apparent transfer of these

systematic orientation effects to the "new" shapes). Alternatively, if subjects were simply resorting to orientation-dependent mechanisms without the benefit of multiple view-based representations, response times would be systematically related only to a single, "canonical" orientation (most likely the upright). However, as shown in the bottom panel, response times exhibit a linear function not with distance from the upright, but with distance from the nearest familiar orientation. Thus, these results indicate that humans may routinely encode both object-based and view-based representations, and that it is the perceptual discrimination task, rather than the training condition, that determines how such representations are ultimately applied.

Several research directions related to these results have been initiated. A pair of control studies have been designed to examine whether the effects of orientation observed for the familiar shapes may result from either of two factors included in the test condition. First, it is possible that the introduction of a wider range of orientations, regardless of stimulus context, is sufficient to shift subjects to an orientation-dependent recognition strategy. Therefore, one control experiment uses an identical practice phase, but simply introduces a wider range of orientations without the inclusion of new shapes. Likewise, it is possible the introduction of new shapes, regardless of whether they are confusable with the original shapes, is also sufficient to shift subjects' strategy. Therefore, another control experiment uses an identical practice phase, but introduces a wider range of orientations with new shapes that do not share a significant number of parts with the old shapes. Additionally, an experiment has been designed to demonstrate that subjects' recognition strategy may be shifted in the opposite direction: here subjects will initially learn and recognize all of the stimuli, both old and new shapes, the contrasts between them presumably leading subjects to adopt the use of an orientation-dependent recognition mechanism. However, the test condition will remove three of the shapes so as allow the remaining shapes to be differentiated along a single dimension. Here, if a shift to orientation-independent mechanisms is made, the introduction of additional orientations should not produce systematic variations in response times — confirming that it is the recognition context and not the learning conditions that determine which mode of representation is used. Finally, we have begun to extend these studies to simple 3D objects — the objective being to demonstrate the generality of both object-based and view-based representations.

5. Object Configurations and Recognition Mechanisms

As discussed in the previous section, Tarr and Pinker (1990), have demonstrated that stimulus context, in particular, the relationship between the configurations of individual stimulus items, has consequences for whether recognition is orientation-dependent or orientation-independent. Accordingly, it should be possible to design a stimulus set for which all shapes are matched for complexity and symmetry, and the recognition strategy for subsets of such shapes is determined by which shapes are included in the recognition subset — according to the number of dimensions that must be attended to in order to differentiate between them. After developing such a set, this has been investigated by selecting subsets in which shapes differ: along a single dimension ("1D" condition, one such subset shown in Figure 7); along one dimension, plus inboard and outboard position along that dimension ("1D+" condition, one such subset shown in Figure 8); or, along two dimensions ("2D" condition). Each subset were used in a simple recognition paradigm across many orientations. Because Tarr and Pinker hypothesized that object-based frames may only distinguish between the ordering of parts along a single dimension, as well as the inboard/outboard positions of

parts adjacent to that axis, it was predicted that the first subset would produce the strongest evidence for the use of orientation-independent mechanisms, the second subset would produce similar results (although with some hints of orientation-dependent effects due to the possible mixing of strategies by individual subjects or differences in strategies between subjects), and the third subset would produce clear evidence for orientation-dependent mechanisms.

As illustrated in the upper panel of Figure 9, the results for the 1D condition are consistent with previous results. In particular, even in the initial phases of the experiment only small effects of orientation are observed. Such effects are outside the range normally considered as signatures of a mental transformation, rather they are explained as orientation-dependent preprocessing mechanisms, for instance, locating the major axes of the shape (Tarr & Pinker, 1991, for a discussion of these issues). Indeed, even these small effects of orientation diminish with practice (Figure 10) — possibly because subjects automaticize the preprocessing of shapes when they are familiar, or because occasional trials in which an orientation-dependent recognition mechanism is used diminish with familiarity. Moreover, the majority of the initial effects of orientation arise from a single subset of stimuli — other subsets do not yield even small effects. Thus, it is possible that an idiosyncratic contrast within this particular subset prompted subjects to rely to a greater extent on orientation-dependent mechanisms — an issue that we are currently investigating. Similar effects are observed for the 1D+ condition. As illustrated in the bottom panel of Figure 9, response times are again only marginally influenced by the orientations of the shapes and this effect diminishes with practice (Figure 10). While the overall measured putative rate of rotation is closer to that associated with orientation-dependent recognition (Figure 10), it is possible that the added contrast of inboard/outboard relations prompted some subjects to adopt an orientation-dependent recognition mechanism. Thus, it is not surprising that slightly greater effects of orientation are observed in this condition. Finally, the 2D condition is currently being run — with results predicted to exhibit a clear effect of orientation consistent with the magnitudes found in studies where the use of orientation-dependent recognition mechanisms is unambiguous.

6. Lexical and Perceptual Encoding of Spatial Relations

In conjunction with William Hayward, a Ph.D. student at Yale, I have begun a series of experiments to investigate the nature of *qualitative* spatial relations encoded between object parts, objects, or places. Our hypothesis (and one also held by others that has recently generated a great deal of interest) is that the restricted meanings of spatial prepositions used in language reflect the underlying structure of our visual representation system. Therefore, the study of the structure of spatial terms may be used a "backdoor" to elucidating spatial relations coded in vision (at present, while many theories posit such relations, e.g., Biederman's, 1987, RBC, no one has systematically cataloged or characterized these relations). Ultimately, such concepts may provide one of the basic ways of moving between an inherently spatial representation and a linguistic description. We have started with a series of experiments designed to illuminate the constraints implicit in spatial terms, such as "above" and "below," as well as their relationship to purely perceptual representations, such as object and scene representations.

The project involved two distinct phases. The first, the results of which are summarized in Figures 11 and 12, had subjects consider the spatial configurations that are captured by spatial prepositions such as "above." Two objects were presented in a

scene where one, the reference object, always appeared in the center (depicted as a computer in the figures), and the other, the figural object, appeared in one of many positions on a 7x7 grid surrounding the reference object. In the first experiment, subjects were simply asked to apply the term they felt was most appropriate for the spatial relationship observed between the two objects. Figure 11 illustrates the tabulated frequencies of usage of vertically oriented, e.g., "above" and "below," and horizontally oriented, e.g., "left" and "right," spatial prepositions. In the second experiment, subjects were given the spatial preposition and asked to rate its applicability to the depicted spatial configuration. Figure 12 illustrates subjects' ratings of applicability for both vertically and horizontally oriented prepositions with the location of the figural object on the grid. Results from both experiments suggest that subjects have a preference to use these spatial terms in a specific manner — when the figural object is directly vertical or directly horizontal relative to the referent. Moreover, while subjects will sometimes use the same spatial prepositions to describe other spatial relations, they do so in a gradient that decreases in both frequency of use and assessed appropriateness.

The second phase involved perceptual judgments using objects configured as in the first phase. The first experiment required subjects to assess and then describe the spatial relationship between the two objects and then recall the position of the figural object by indicating the center of its mass with a pointing device. Figure 13 illustrates subjects' precision in estimating location as separate function in the horizontal and vertical directions. The distortion horizontally for each estimate is shown in (a) and vertically in part (b). Subjects are far more accurate in estimating the horizontal position of the figural object when it is in a location described as above or below in the previous experiments; likewise, subjects are more accurate in estimating the vertical position of the figural object when it is in a location described as left or right in the previous experiment. Surprisingly, the identical task, but without the preceding verbal description of the spatial relationship, yielded nearly uniform errors in estimating the position of the figural object across all of its locations. Thus, it appears that to some extent, the qualitative coding of spatial relations is a *lexicalization effect* — in that, systematic distortions towards horizontal or vertical occur only when subjects are required to categorize the spatial relationship for purposes of encoding it lexically. Such a finding would have important implications for how humans encode spatial relations both between and within objects, suggesting that object and scene representations are actually more similar to the originally perceived spatial layout, retaining specific point-by-point, rather than qualitative, relationships between objects or parts.

However, it is possible that this experiment simply masks effects of perceptually-based qualitative encoding. In particular, the large errors introduced by having subjects recall the absolute position of the figural object may have obscured any systematic effects in the no-description condition — hence the apparently uniform error fields. To address this possibility, the no-description condition was rerun with a relative position judgment in which subjects simply judged whether the figural object was in the same location relative to the reference object in two sequential frames (which shifted randomly in screen position so that subjects could not simply note the absolute position of the figural object between frames). Figure 14 illustrates that this manipulation did reveal qualitative effects within a purely perceptual judgment. In particular, sensitivity for discriminating position was markedly higher at locations where the figural object was vertically or horizontally aligned with the reference object. Thus, a strong version of the lexicalization hypothesis seems unlikely — here a task that did not invoke lexical structures still resulted in systematic distortions in subjects' perceptual memory —

rather, there is apparently at least weak qualitative encoding at the perceptual level, possibly significant enough to explain the categorical effects found in phase one as non-lexical (in particular, because distortions here occurred at the same regions adjudged to captured best by spatial prepositions in earlier experiments). Of course, the large lexicalization effect found in the previous experiment suggests otherwise, instead making it likely that there is an interaction between spatial relations as captured by the visual and linguistic systems.

These results suggest that there is a close connection between inherently spatial representations of the world, for instance those found in the visual system, and the categorical form referred to in language (i.e., we refer to objects being simply *above* rather than precisely how far above). Given these basic finding, studies must be developed to address the precise nature of that relationship. Interestingly, it has been suggested that our use of language constrains the spatial relationships that we visually represent — a hypothesis we found some evidence for — the conditions under which this is true, as well as exactly what constraints are imposed by linguistic structures, may provide a better understanding of object and scene representations that rely on qualitatively encoded spatial or structural relations. Conversely, it is likely that how language structures space has been strongly influenced by our visual representation system — how such constraints manifest themselves is crucial for understanding the underlying topology of our concepts of space.

REFERENCES

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.

Cooper, L. A., & Schacter, D. L. (1992). Dissociations between structural and episodic representations of visual objects. *Current Directions in Psychological Science*, *1*(5), 141-146.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review*, *95*(12), 115-123.

Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain-specificity in visual object recognition. *Current Directions in Psychological Science*, *1*(5), 164-169.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, *13*, 289-303.

Jolicoeur, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind & Language*, *5*(4), 387-410.

Jolicoeur, P., & Milliken, B. (1989). Identification of disoriented objects: Effects of context of prior presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 200-210.

Koenderink, J. J. (1987). An internal representation for solid shape based on the topological properties of the apparent contour. In W. Richards & S. Ullman (Eds.), *Image Understanding: 1985-86* (pp. 257-285). Norwood, NJ: Ablex Publishing Corporation.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: Freeman.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501-518.

Schacter, D. L., Cooper, L. A., Delaney, S. M., Peterson, M. A., & Tharan, M. (1991). Implicit memory for possible and impossible objects: Constraints on the construction of structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(1), 3-19.

Schacter, D. L., Cooper, L. A., & Delany, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General, 119,* 5-24.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Tarr, M. J. (1990). A different perspective on human object recognition: The identification of objects in unfamiliar views. Presented at *The 31st Annual Meeting of the Psychonomic Society*, November 16-18. New Orleans, LA.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*(28), 233-282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science, 1*(42), 253-256.

Tarr, M. J., & Pinker, S. (1991). Orientation-dependent mechanisms in shape recognition: further issues. *Psychological Science, 2*(32), 207-209.

PUBLICATIONS

Tarr, M. J. (1992). A context for AI? *IEEE Expert.*, **7**(6), 88-89. Book Review.

Tarr, M. J., & Black, M. J. (In Press). A computational and evolutionary perspective on the role of representation in vision. *Computer Vision, Graphics, and Image Processing: Image Understanding.*

Tarr, M. J. (In Press). From Perception to Cognition. Commentary to appear in *Behavioral and Brain Sciences.*

Tarr, M. J. (In Press). Is a picture really worth a thousand words? To appear in *Computational Intelligence.*

Tarr, M. J. (Under Review). Behavioral and computational constraints in human object representation. Submitted to *Psychological Bulletin.*

Tarr, M. J., & Kriegman, D. J. (In Preparation). A formal basis for understanding view-based representations in humans. To appear in book of papers from the *Workshop on Visual Perception: Computation and Psychophysics.*

Tarr, M. J. (In Preparation). Visual representation. To appear in *Encyclopedia of Human Behavior.* San Diego, CA: Academic Press.

Neath, I., & Tarr, M. J. (Under Revision). Splitting the spotlight: The allocation of visual attention across noncontiguous spatial locations in cued luminance-detection tasks. To be resubmitted to *Journal of Experimental Psychology: Human Perception and Performance.*

PERSONNEL

Doctoral Students	William Hayward	2nd Year
	Alan Ashworth	3rd Year
Undergraduates	Steven Messé	BA in May 1993
	Preston Bost	""
	Douglas Bitting	""

CONFERENCE PRESENTATIONS AND INVITED COLLOQUIA

Tarr, M. J. Behavioral and computational constraints in human shape representation. Presented at the *Conference on Cognition and Representation*, State University of New York at Buffalo, April 3-5, 1992.

Tarr, M. J., & Black, M. J. Psychophysical implications of temporal persistence in early vision: A computational account of representational momentum. Presented at the *Annual Meeting of The Association for Research in Vision and Ophthalmology (ARVO)*, Sarasota, Florida, May 3-8, 1992.

Tarr, M. J., & Kriegman, D. J. Viewpoint-dependent image features in human object representation. Presented at the *33rd Annual Meeting of the Psychonomic Society*, St. Louis, November 13-15, 1992.

Tarr, M. J., & Kriegman, D. J. A formal basis for understanding view-based representations in humans. *Workshop on Visual Perception: Computation and Psychophysics*, Cape Cod, MA, January, 1993.

Tarr, M. J. Invited panel member, special session on purposive vision. *International Joint Conference on Artificial Intelligence*, Chambery, France, August, 1993.

COLLOQUIA: IBM Watson Research Laboratory, February, 1992; GRASP Laboratory, University of Pennsylvania, February 1992; Department of Psychology, Cornell University, April, 1992; Department of Psychology, Carnegie-Mellon University, November, 1992; Robotics Institute, Carnegie-Mellon University, November, 1992; Haskins Laboratory, November, 1992; Department of Cognitive and Neural Systems, Boston University, February, 1993; Department of Psychology, University of Toronto, March, 1993; Department of Cognitive Science, Brown University, April, 1993.

OTHER

PRODUCT DONATION: *Viewpoint Animation Engineering, Inc.* Donation of extensive collection of three-dimensional models depicting many natural objects for CAD-based solid modeling systems. To be used in psychophysical experiments of human object recognition. Approximate retail value: $7,000.

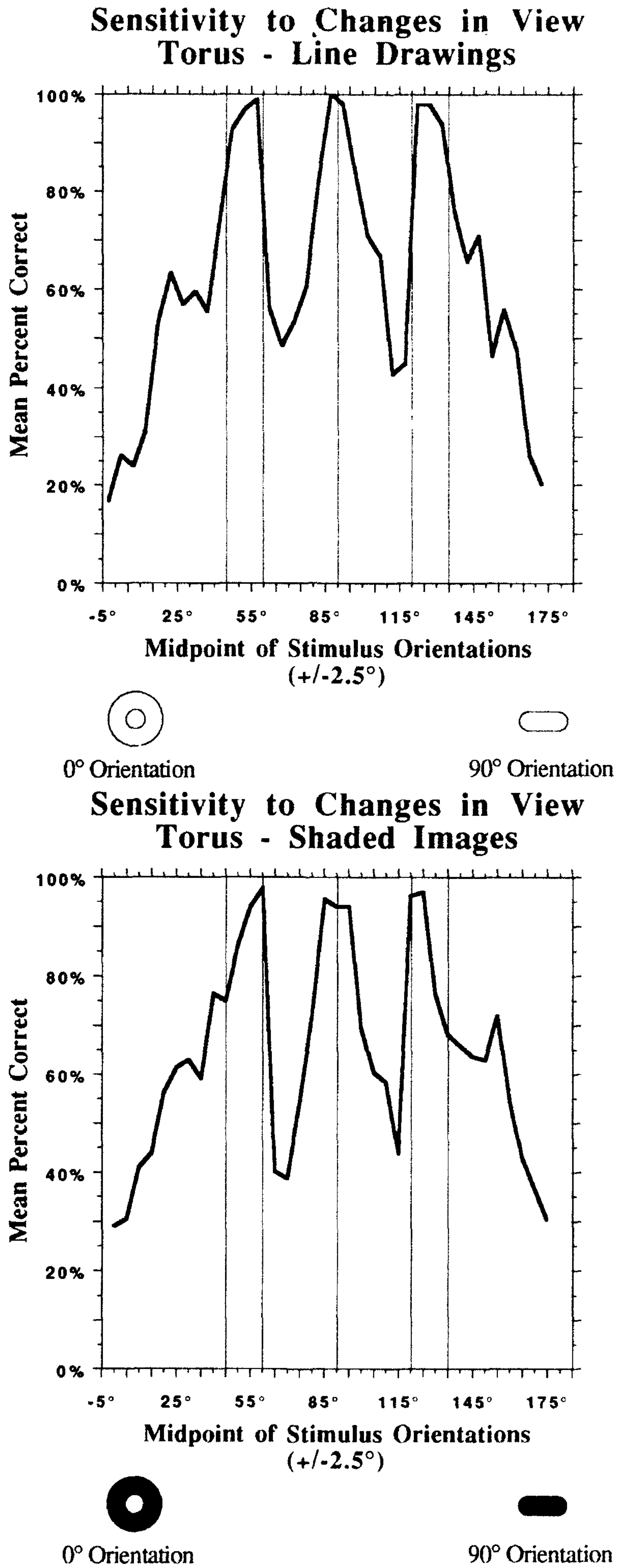
Appointed as a consulting editor, *Journal of Experimental Psychology: Human Perception and Performance.*

Technical advisory board. *Advanced Robotic Technologies, Inc.*

Table 1.

Object	Canonical Orientation	Mean Rating
1. Airplane	-45°	5.8776
2. Apple	0°	6.5510
3. Axe	90°	5.1633
4. Baking Pan	-135°	3.6327
5. Banana	45°	5.3265
6. Bike	90°	6.4286
7. Blender	0°	6.2653
8. Bolt	-90°	6.0816
9. Camera	-45°	5.6122
10. Candle	0°	6.5306
11. Chisel	0°	5.8163
12. Clothespin	90°	5.8776
13. Cow	0°	----
14. Desk Chair	0°	6.0612
15. Dining Table	0°	5.4694
16. Dinner Plate	0°	4.2449
17. Dumbbell	0°	6.3469
18. End Table	135°	4.3673
19. Fire Extinguisher	0°	6.0204
20. Flashlight	45°	5.6531
21. Flower	-45°	5.8571
22. Frisbee	180°	3.2041
23. Glass	0°	6.4490
24. Claw Hammer	-90°	6.3469
25. Lamp	0°	6.7551
26. Lemon	0°	6.0611
27. Light Bulb	0°	6.4898
28. Macintosh Computer	0°	6.5306
29. Office Desk	0°	5.9592
30. Padlock	0°	6.3469
31. Peach	-45°	4.5306
32. Pear	0°	6.6735
33. Pencil	0°	6.6939
34. Phone	0°	6.5306
35. Pipe	90°	6.1837
36. Pocket Watch	0°	3.8163
37. Pumpkin	-135°	6.3265
38. Screwdriver	45°	5.9592
39. Stuffed Chair	0°	5.4288
40. Task Chair	45°	5.6875
41. Tea Cup	0°	4.1837
42. Tea Pot	-90°	5.8980
43. Drinking Glass	45°	4.3265
44. Wood Chair	45°	5.3265

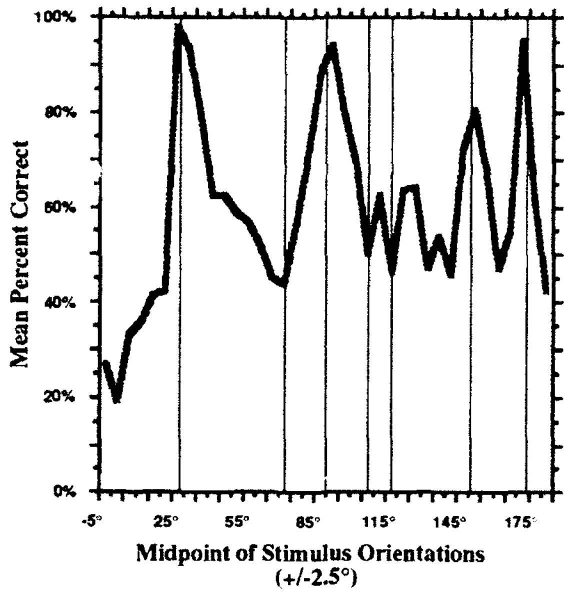
Figure 1.



Sensitivity to Changes in View
Torus - Line Drawings
100%
80%
60%
40%
20%
0%
Mean Percent Correct
-5°
25°
55°
85°
115°
145°
175°
Midpoint of Stimulus Orientations
(+/-2.5°)
0° Orientation
90° Orientation
Sensitivity to Changes in View
Torus - Shaded Images
100%
80%
60%
40%
20%
0%
Mean Percent Correct
-5°
25°
55°
85°
115°
145°
175°
Midpoint of Stimulus Orientations
(+/-2.5°)
0° Orientation
90° Orientation

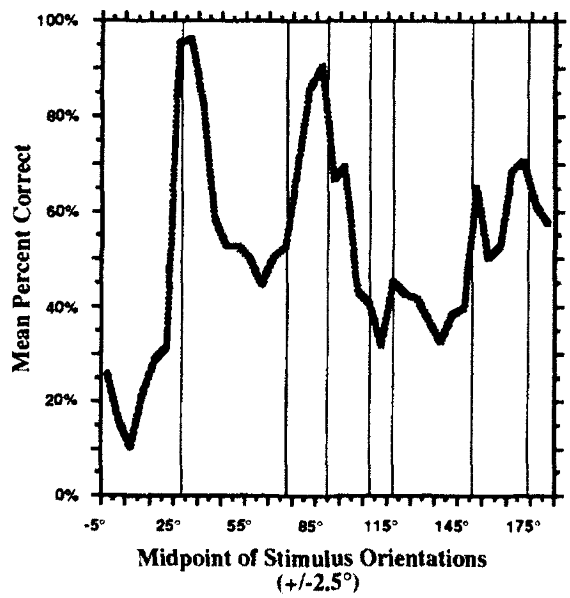
Figure 2.

Bell - Line Drawings



0° Orientation 90° Orientation 180° Orientation

Bell - Shaded Images



0° Orientation 90° Orientation 180° Orientation

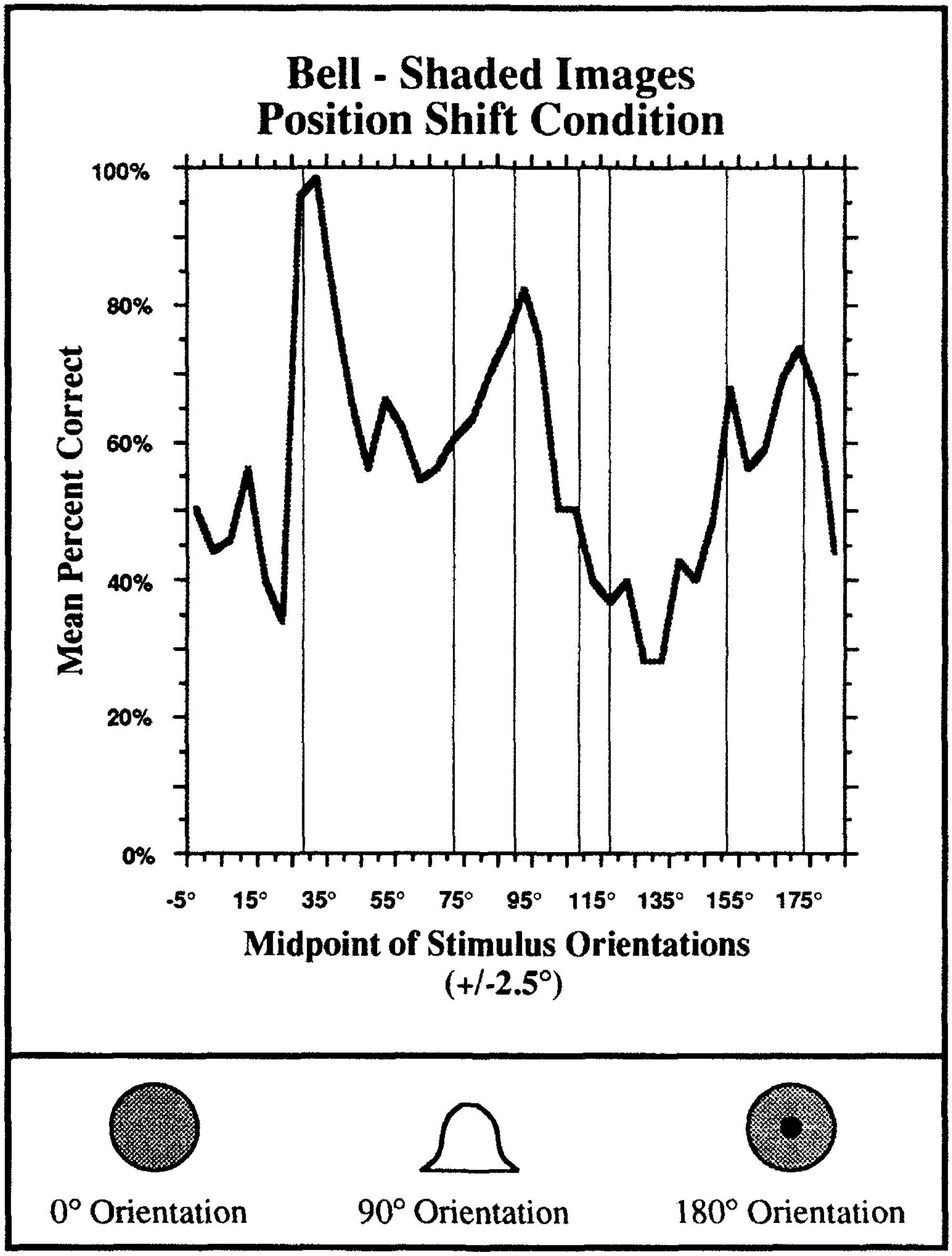
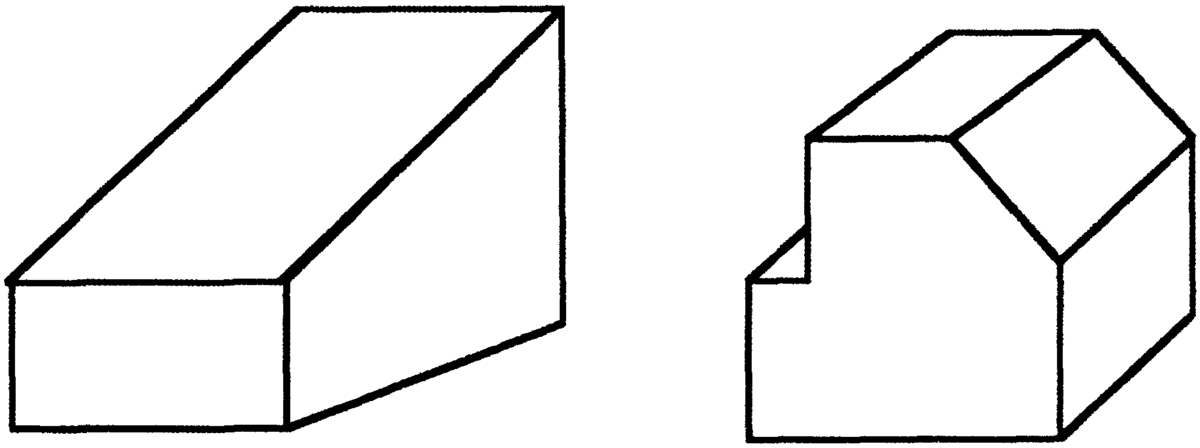
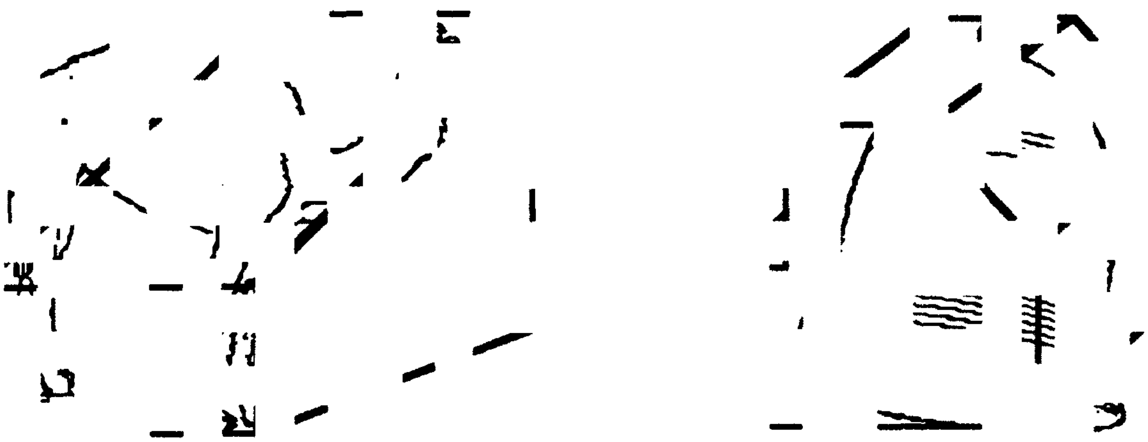

Bell - Shaded Images
Position Shift Condition
Mean Percent Correct
100%
80%
60%
40%
20%
0%
-5°
15°
35°
55°
75°
95°
115°
135°
155°
175°
Midpoint of Stimulus Orientations
(+/-2.5°)
0° Orientation
90° Orientation
180° Orientation

Figure 4.



Prime Objects



Test Displays. Subjects' task is to name the common objects (camel and bell, respectively). Overlapping prime and common objects are presented at successively decreasing levels of degradation — level at which subject correctly names the object is recorded as the response.

Figure 5.

Putative Rate of Rotation Throughout Practice Phase

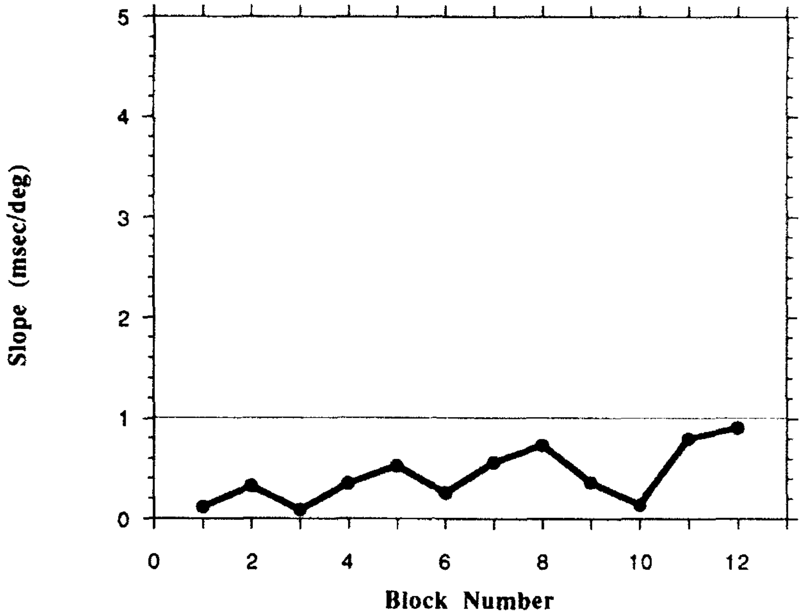
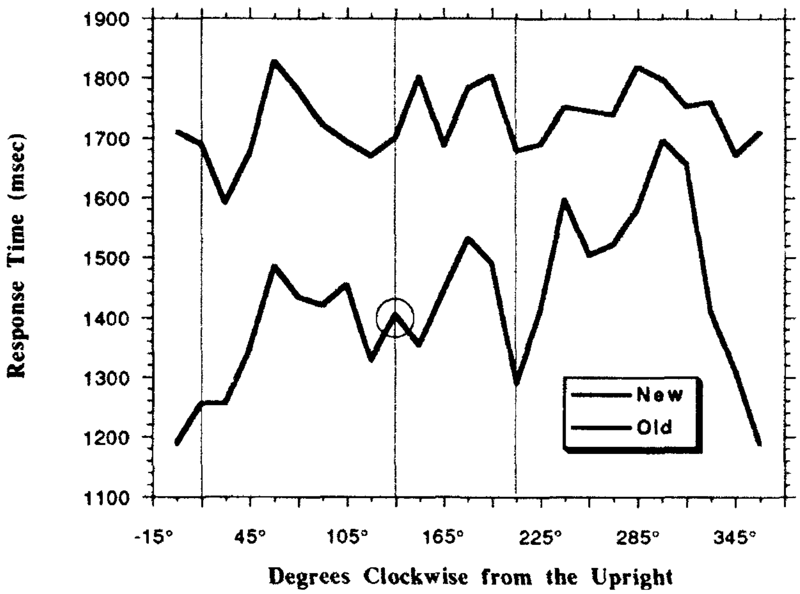


Figure 6.

Recognition of Familiar Objects: Novel Orientations and Novel Distractors



Responses by Distance from Nearest Familiar Orientation

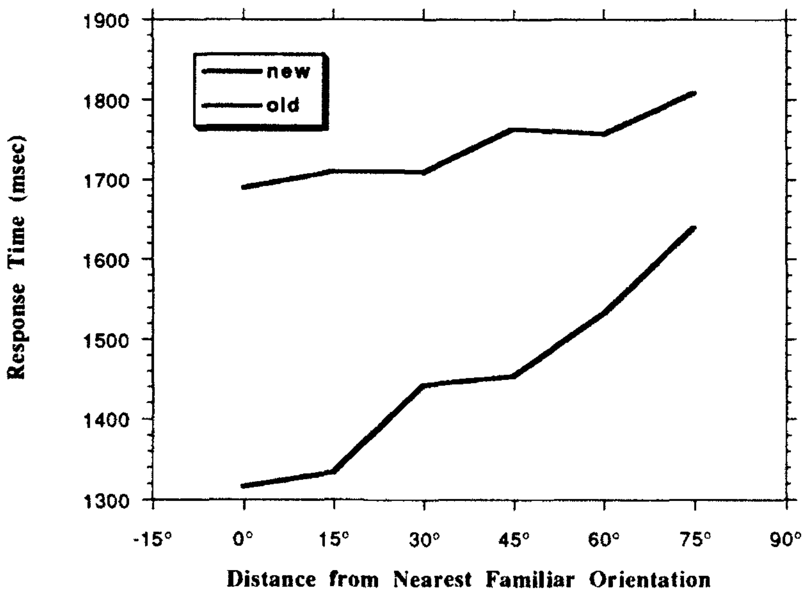


Figure 7.

STIMULUS SHAPES:
DIFFERENTIATED BY LINEAR ORDERING

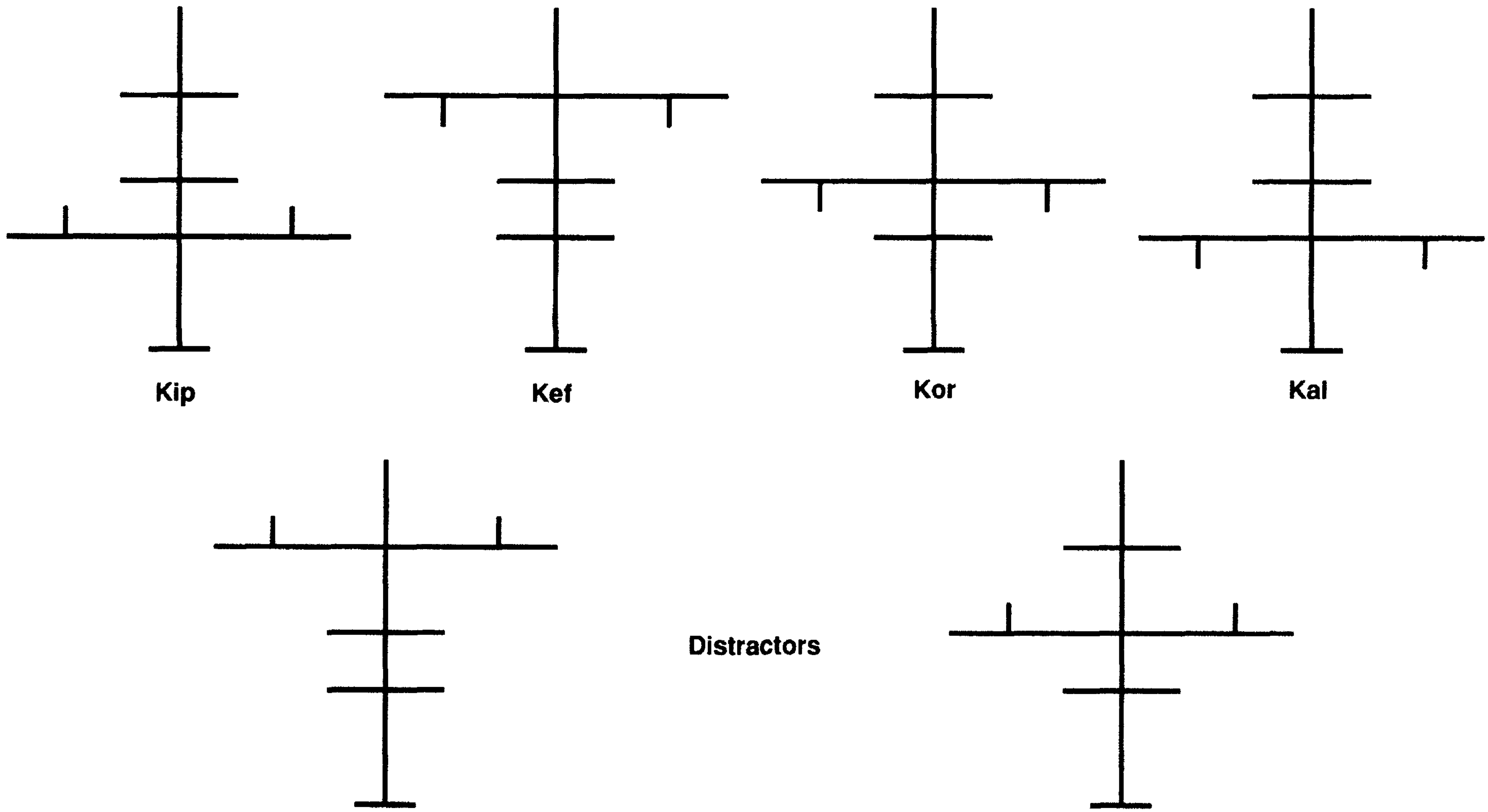


Figure 8.

STIMULUS SHAPES: DIFFERENTIATED BY LINEAR ORDERING + INBOARD/OUTBOARD POSITIONS

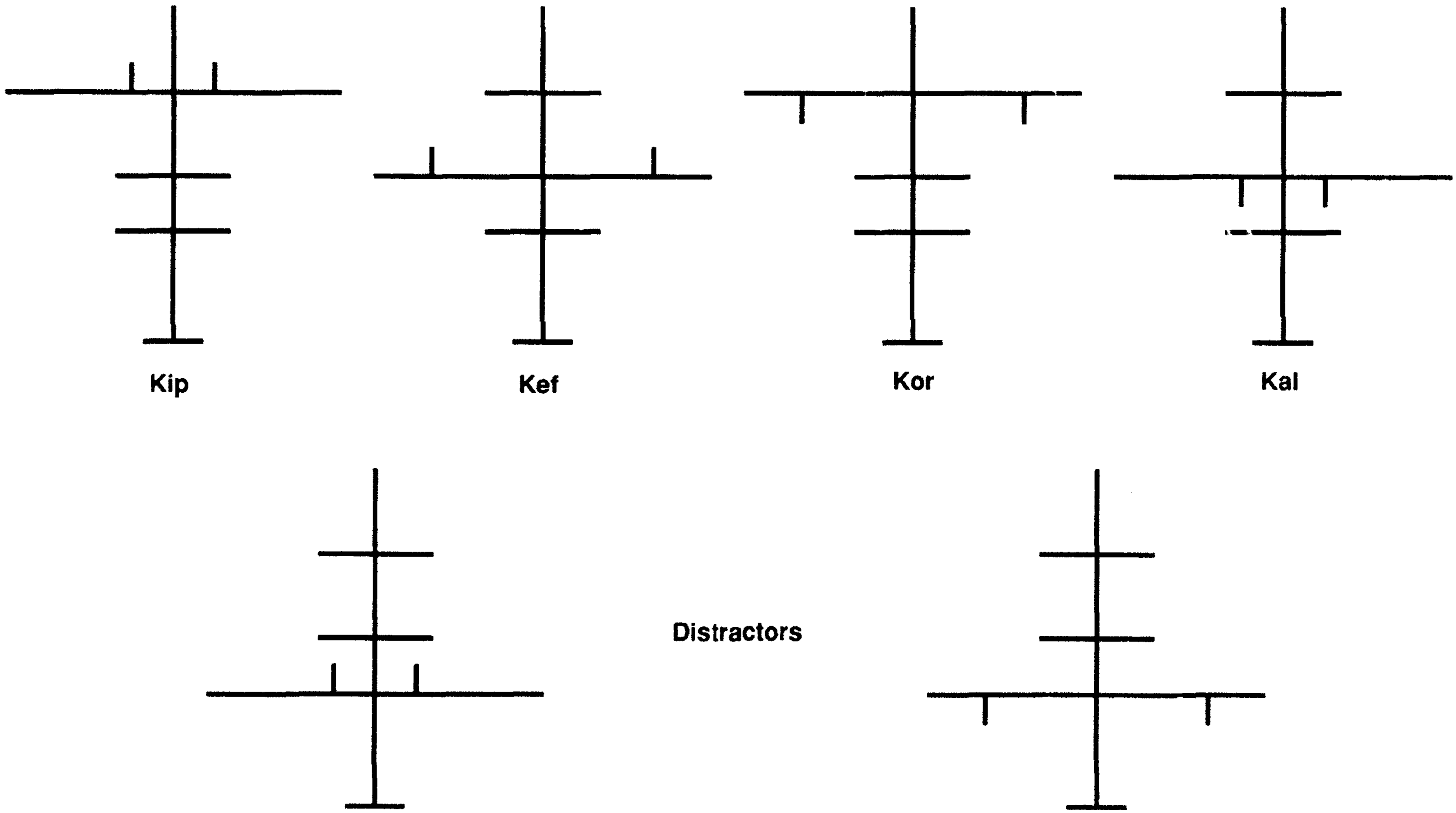
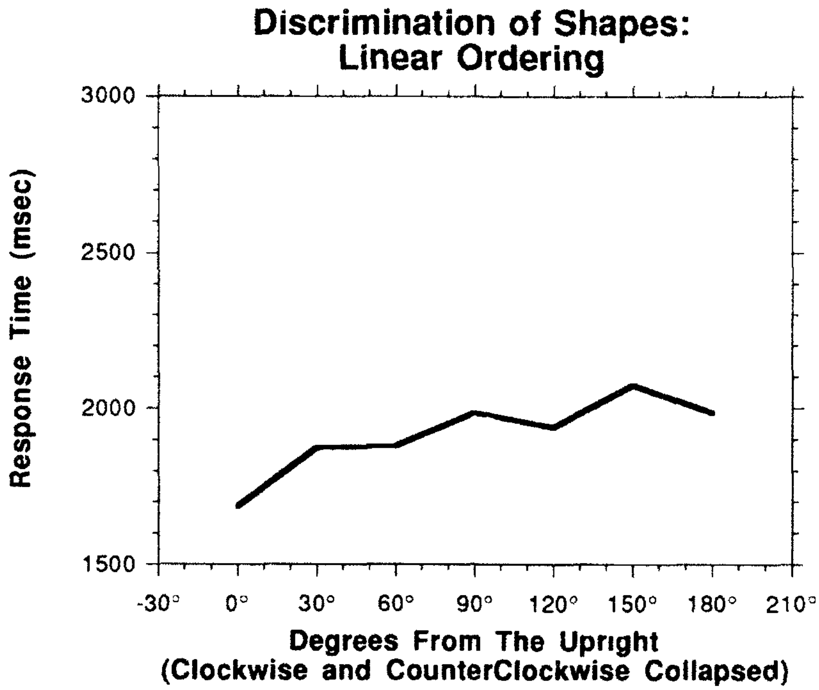
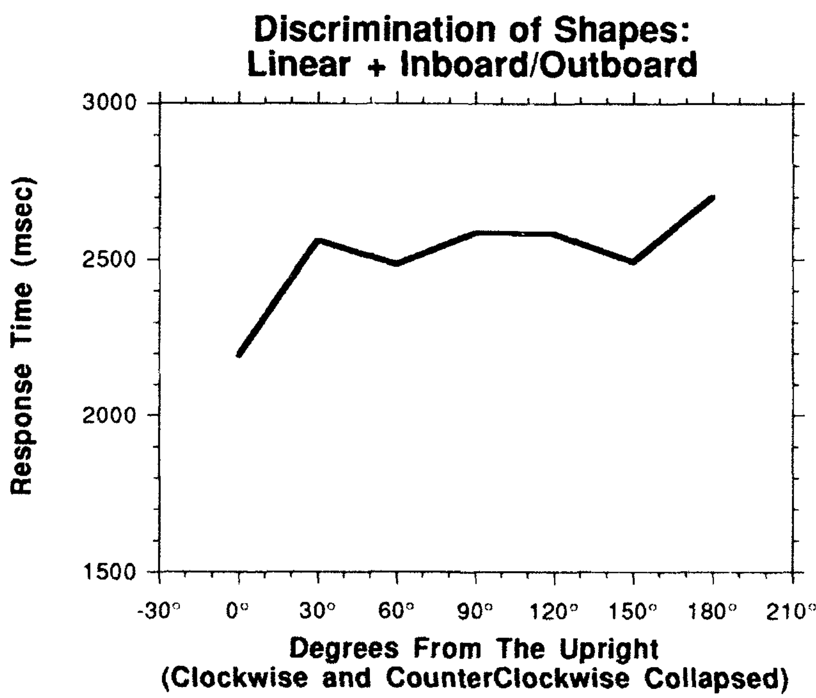


Figure 9.

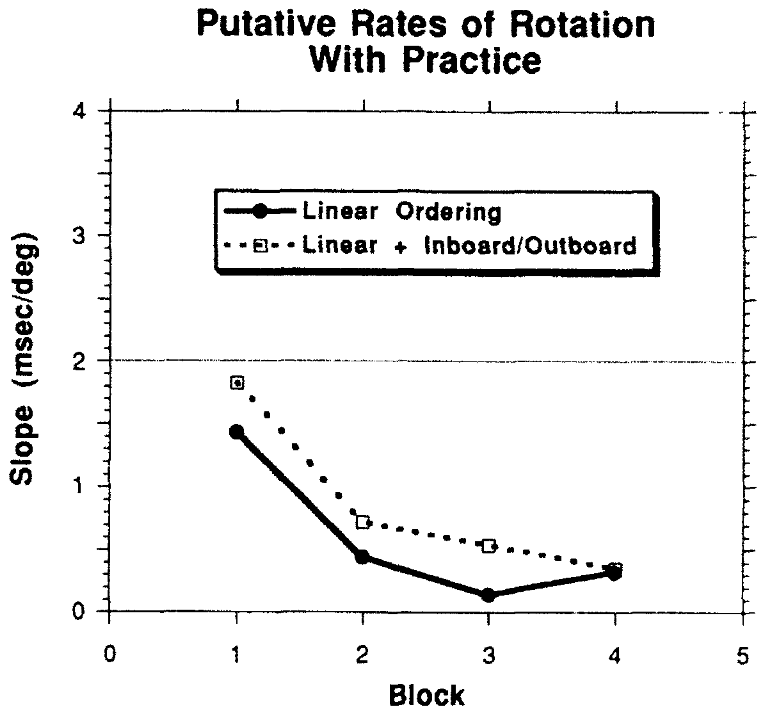


Discrimination of Shapes:
Linear Ordering
3000
2500
2000
1500
Response Time (msec)
-30°
0°
30°
60°
90°
120°
150°
180°
210°
Degrees From The Upright
(Clockwise and CounterClockwise Collapsed)



Discrimination of Shapes:
Linear + Inboard/Outboard
3000
2500
2000
1500
Response Time (msec)
-30°
0°
30°
60°
90°
120°
150°
180°
210°
Degrees From The Upright
(Clockwise and CounterClockwise Collapsed)

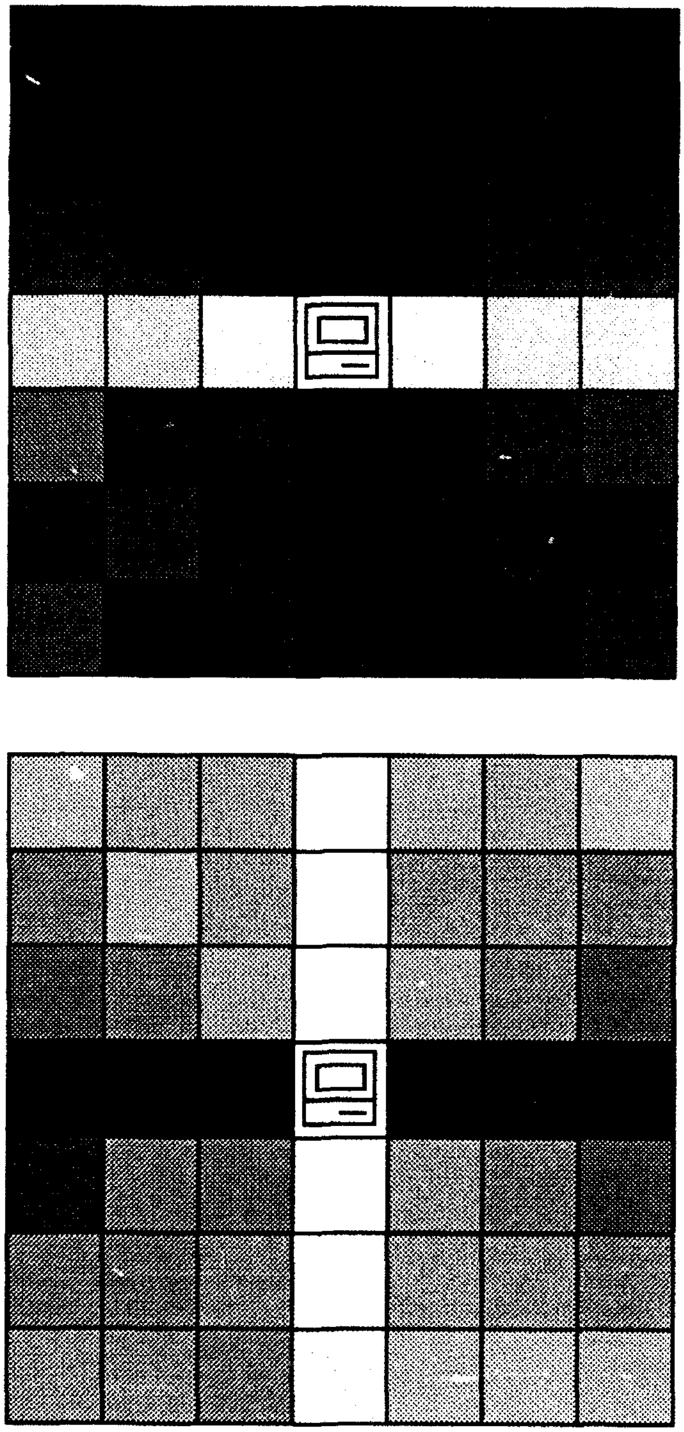
Figure 10.



Putative Rates of Rotation
With Practice
Linear Ordering
Linear + Inboard/Outboard
Slope (msec/deg)
0
1
2
3
4
0
1
2
3
4
5
Block

Figure 11.

Subject Generated Spatial Terms for Perceived Spatial Relationships



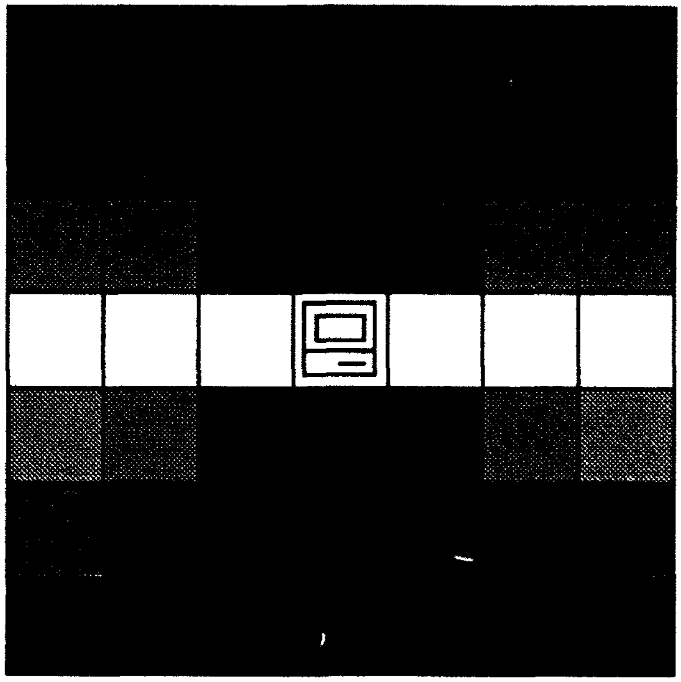
a.

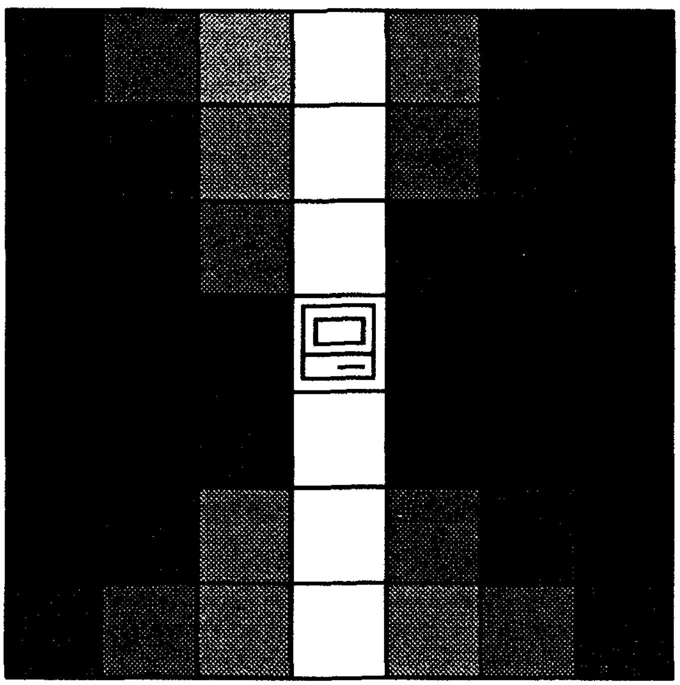
b.

Frequency of use for spatial prepositions describing (a) vertical and (b) horizontal relation between figural object and reference object. Darker areas represent greater frequency of prepositions when the figural object is in that position relative to the reference object.

Figure 12.

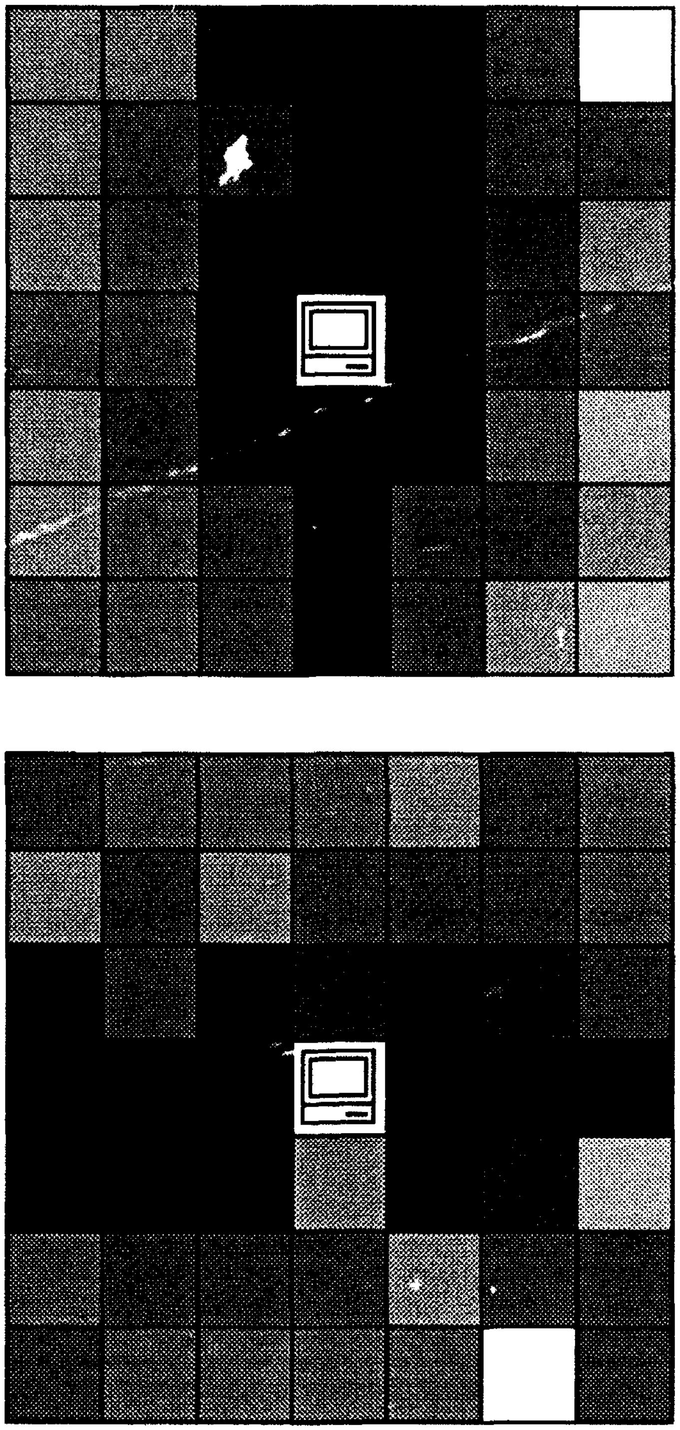
Appropriateness Ratings of Spatial Terms for Perceived Spatial Relationships

a. 

b. 

Ratings of appropriateness of (a) ABOVE/BELOW and (b) LEFT/RIGHT in describing the relationship between a figural object and a reference object. Darker areas represent positions in which the appropriate preposition was judged to describe the relationship well.

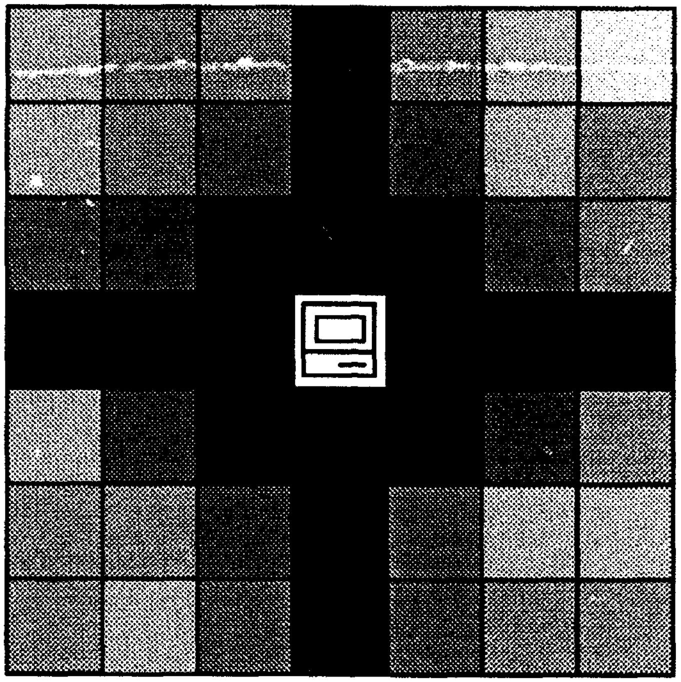
Memory for Position Subsequent to a Lexical Description of the Spatial Relationship



a. b.

Accuracy of estimation of position of figural object in relation to reference object along (a) the horizontal axis and (b) the vertical axis. Darker areas represent greater accuracy. Subjects are more accurate along the horizontal when the figural object is directly vertical of the reference object, and are more accurate along the vertical when the figural object is directly horizontal.

Sensitivity to Shifts in Position Relative to a Reference Object



"Space field" derived from same/different judgments on successive presentations of a figural object. Shifts in the position of the figural object between successive presentations were significantly smaller than the shifts in position of the object from one trial to the next. Darker areas are the positions of the figural object relative to the reference object (shown in the center) that represent positions of greater accuracy. Subjects are more accurate at successive 90° angles aligned or perpendicular to the gravitational upright.